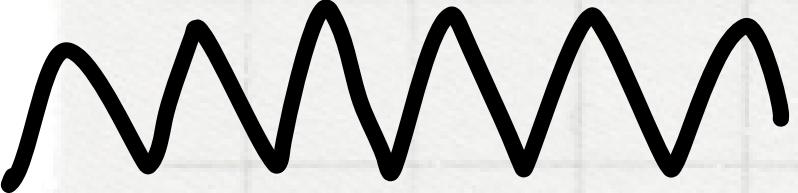


Task 5

Kimia Farma Big Data Analytics Project Based
Internship Program X Rakamin Academy

By Anggita Ramadhani

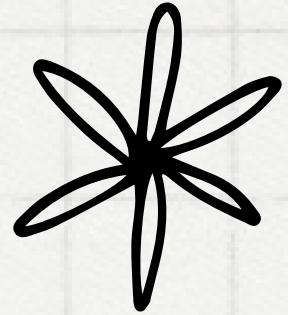


Soal 1

Dari 2 query ini, mana yang bekerja lebih baik? Jelaskan mengapa.

SELECT * FROM pelanggan WHERE SUBSTR(alamat, 1, 3) = Mat;

SELECT * FROM pelanggan WHERE alamat LIKE 'Mat%'



Jawaban : SELECT * FROM pelanggan WHERE alamat LIKE 'Mat%';

Alasan :

- 1.Kurang tanda petik di query pertama, karena Mat merupakan string.
- 2.Query kedua lebih sesuai dengan kebutuhan, yaitu mencari baris di dalam kolom 'alamat' yang di awali dengan 'Mat'. Fungsi 'SUBSTR' pada query pertama cenderung lebih kompleks dan kurang optimal untuk kasus ini.
- 3.Pada query kedua, penggunaan 'LIKE 'Mat%' lebih efisien untuk mencari nilai yang dimulai dengan "Mat" karena dapat memanfaatkan indeks. Sedangkan penggunaan fungsi 'SUBSTR' pada query pertama mungkin dapat memperlambat kinerja, terutama jika tabel memiliki banyak baris.

Soal 2

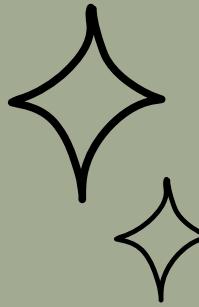
Anggap kita memiliki tabel pelanggan dengan kolom: id, nama, tanggal_lahir, alamat. Bagaimana cara yang lebih tepat dalam menulis query untuk mendapatkan data pelanggan yang tanggal_lahir nya ada di antara 2000-01-01 sampai 2008-12-31? Pilihlah salah satu jawaban dan berikan alasannya.

- a. SELECT * FROM pelanggan WHERE tanggal_lahir >= '2000-01-01' AND tanggal_lahir <= '2008-12-31'
- b. SELECT * FROM pelanggan WHERE tanggal_lahir BETWEEN '2000-01-01' AND '2008-12-31'

Jawaban : SELECT * FROM pelanggan WHERE tanggal_lahir BETWEEN '2000-01-01' AND '2008-12-31'

Alasan :

- 1.Query menggunakan 'BETWEEN' lebih sederhana dan mudah dipahami.
- 2.Penggunaan 'BETWEEN' memberi ekspresi yang lebih jelas dan intuitif untuk mencari data dalam rentang tertentu.



Soal 3

Tentukan primary key dari table penjualan.

Jawaban : id_invoice

Alasan :

1. id_invoice merupakan kolom yang unik dan mengidentifikasi setiap transaksi penjualan secara eksklusif.
2. Tidak mempunyai nilai NULL, karena primary key tidak dapat memiliki nilai NULL.

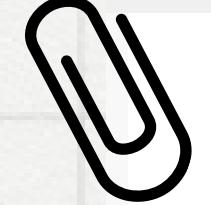
A	B	C	D	E	F	G	H	I	J	K	L
id_distributo	id_cabang	id_invoic	tanggal	id_customer	id_barang	jumlah_b	unit	harga	mata_uai	brand_id	lini
TD	CAB01	IN5997	20/01/2022	CUST55380	BRG0001	1	DUS	1169,91	IDR	BRND001	OGB & F
TD	CAB01	IN6297	20/01/2022	CUST55381	BRG0002	5	DUS	2337,5	IDR	BRND002	ETIKAL
TA	CAB02	IN6155	21/01/2022	CUST55382	BRG0003	9	DUS	10690,6	IDR	BRND003	MARCKS
EPM	CAB03	IN6144	22/01/2022	CUST55383	BRG0004	13	DUS	8700,7	IDR	BRND004	VNS
TD	CAB04	IN6280	23/01/2022	CUST55384	BRG0005	1	DUS	5648,3	IDR	BRND005	SLCYL
TA	CAB05	IN6052	23/01/2022	CUST55385	BRG0006	5	DUS	2819,2	IDR	BRND006	OGB & F
EPM	CAB06	IN6089	25/01/2022	CUST55386	BRG0007	9	DUS	4592,1	IDR	BRND007	ETIKAL
TD	CAB07	IN6251	23/01/2022	CUST55387	BRG0008	1	DUS	3991,9	IDR	BRND008	MARCKS
TA	CAB08	IN6036	27/01/2022	CUST55388	BRG0009	23	DUS	5780,7	IDR	BRND009	VNS
EPM	CAB09	IN6029	28/01/2022	CUST55389	BRG0010	4	DUS	6940,3	IDR	BRND010	SLCYL
TD	CAB10	IN6190	23/01/2022	CUST55390	BRG0001	2	DUS	1169,91	IDR	BRND001	OGB & F
TD	CAB01	IN6287	30/01/2022	CUST55391	BRG0002	17	DUS	2337,5	IDR	BRND002	ETIKAL
TA	CAB01	IN6102	31/01/2022	CUST55392	BRG0003	11	DUS	10690,6	IDR	BRND003	MARCKS
EPM	CAB02	IN6301	01/02/2022	CUST55393	BRG0004	12	DUS	8700,7	IDR	BRND004	VNS
TD	CAB03	IN6024	27/01/2022	CUST55394	BRG0005	12	DUS	5648,3	IDR	BRND005	SLCYL
TA	CAB04	IN6184	28/01/2022	CUST55395	BRG0006	25	DUS	2819,2	IDR	BRND006	OGB & F
EPM	CAB05	IN6307	23/01/2022	CUST55396	BRG0007	14	DUS	4592,1	IDR	BRND007	ETIKAL
TD	CAB06	IN6245	30/01/2022	CUST55397	BRG0008	34	DUS	3991,9	IDR	BRND008	MARCKS
TA	CAB07	IN6195	31/01/2022	CUST55398	BRG0009	67	DUS	5780,7	IDR	BRND009	VNS
EPM	CAB08	IN6023	01/02/2022	CUST55399	BRG0010	15	DUS	6940,3	IDR	BRND010	SLCYL

Soal 4

Buatlah design datamart (Terdiri dari tabel base, dan tabel aggregate). Upload file query dalam gdrive mu (pastikan dapat diakses public). Lalu masukkan linknya di tabel di bawah, dan cantumkan juga screenshot query nya (jika lebih dari 1 file, maka masing masing file di-screenshot)

No	Nama File	Link
1	create_database.sql	Click here
2	create_base_table.sql	Click here
3	create_aggregat_table.sql	Click here

Create base_table



```
-- Membuat base table
CREATE TABLE base_table AS (
    SELECT
        p.id_invoice,
        p.id_distributor,
        p.id_cabang,
        pl.cabang_sales,
        p.tanggal,
        p.id_customer,
        pl.nama,
        p.id_barang,
        b.nama_barang,
        b.kemasan,
        p.jumlah_barang,
        p.harga,
        b.kode_lini,
        p.lini,
        p.brand_id,
        pl.id_group,
        pl.`group`
    FROM penjualan p
    LEFT JOIN pelanggan pl ON p.id_customer = pl.id_customer
    LEFT JOIN barang b ON p.id_barang = b.kode_barang
);
```

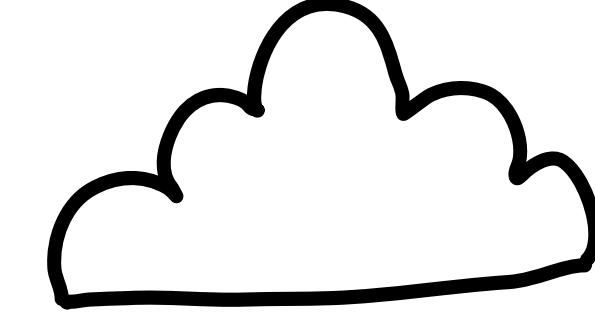


Table Base "base_table"

columns	data type	description	transformation
id_invoice	varchar	ID setiap transaksi penjualan	-
id_distributor	varchar	ID setiap distributor	-
id_cabang	varchar	ID setiap cabang	-
cabang_sales	varchar	lokasi cabang	-
tanggal	date	tanggal transaksi	konversi tipe data menjadi date
id_customer	varchar	ID setiap pelanggan	-

Table Base "base_table"

columns	data type	description	transformation
nama	varchar	nama perusahaan pelanggan	-
id_barang	varchar	ID setiap barang yang terjual	-
nama_barang	varchar	nama produk	-
kemasan	varchar	kemasan untuk setiap produk	-
jumlah_barang	integer	Jumlah produk yang terjual	konversi tipe data menjadi integer
harga	float	Harga satuan pada produk	konversi tipe data menjadi float

Table Base "base_table"

columns	data type	description	transformation
kode_lini	varchar	kode merek	-
lini	varchar	nama merek	-
brand_id	varchar	ID merek	-
id_group	varchar	ID grup	-
group	varchar	kelompok perusahaan	-

④ -- Membuat aggregat table

```
CREATE TABLE aggregat_table AS (
    SELECT
        id_invoice,
        id_distributor,
        id_cabang,
        cabang_sales,
        tanggal,
        MONTHNAME(tanggal) AS bulan,
        id_customer,
        nama AS company_name,
        id_barang,
        nama_barang AS produk,
        kemasan,
        jumlah_barang,
        harga,
        ROUND(SUM(jumlah_barang * harga), 2) AS total,
        kode_lini AS kode_brand,
        lini AS brand,
        brand_id,
        id_group,
        `group` AS group_market
    FROM base_table
    GROUP BY 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19
);
```

Create aggregate_table

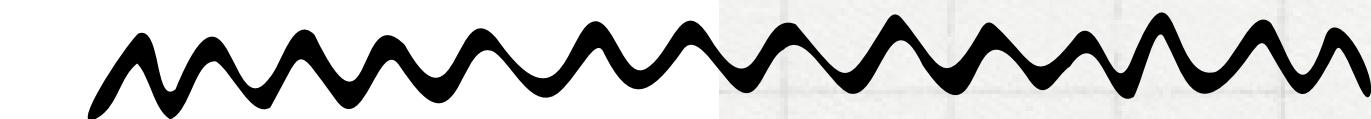
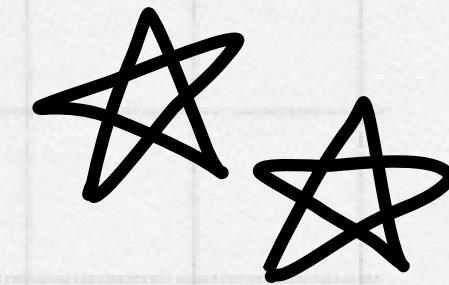


Table Aggregat "aggregat_table"

columns	data type	description	transformation
id_invoice	varchar	ID setiap transaksi penjualan	-
id_distributor	varchar	ID setiap distributor	-
id_cabang	varchar	ID setiap cabang	-
cabang_sales	varchar	lokasi cabang	-
tanggal	date	tanggal transaksi	-
bulan	varchar	bulan transaksi	membuat kolom baru dengan monthname
id_customer	varchar	ID setiap pelanggan	-

Table Aggregat "aggregat_table"

columns	data type	description	transformation
company_name	varchar	nama perusahaan pelanggan	ganti nama kolom jadi company_name
id_barang	varchar	ID setiap barang yang terjual	-
produk	varchar	nama produk	ganti nama kolom jadi produk
kemasan	varchar	kemasan untuk setiap produk	-
jumlah_barang	integer	Jumlah produk yang terjual	-
harga	float	harga satuan pada produk	-

Table Aggregat "aggregat_table"

columns	data type	description	transformation
total	double	nilai harga pada setiap produk dikali dengan jumlah_barang	ROUND(SUM(jumlah_barang * harga), 2)
kode_brand	varchar	kode merek	ganti nama kolom jadi kode_brand
brand	varchar	nama merek	ganti nama kolom jadi brand
brand_id	varchar	ID merek	-
id_group	varchar	ID grup	-
group_market	varchar	kelompok perusahaan	ganti nama kolom jadi group_market

Soal 5

Buatlah data visualisasinya, dan cantumkan linknya di bawah (pastikan bisa diakses publik). Lalu cantumkan juga screenshot visualisasinya

Link Visualisasi :

<https://lookerstudio.google.com/reporting/4b17c3c0-de9f-4ff5-817a-958dcef0705e>

Data Visualization

Sales Report Dashboard

Final Task Project Based Internship Programs

Big Data Analytics - Kimia Farma



2 Jan 2022 - 6 Des 2022

Branch Area

Product

Total Transaction

350

Total Companies

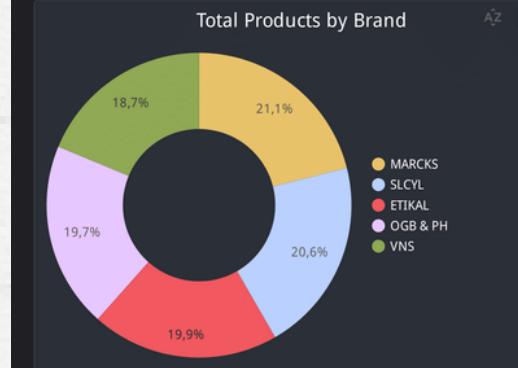
8

Number of Products (Sold)

9.176

Revenue

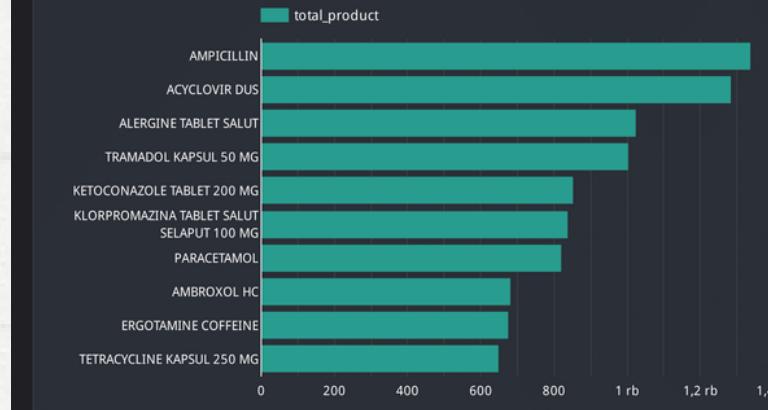
Rp49.216.464,67



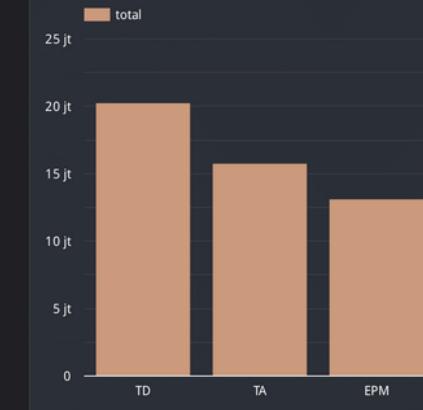
Total Revenue by Month



Sales by Products



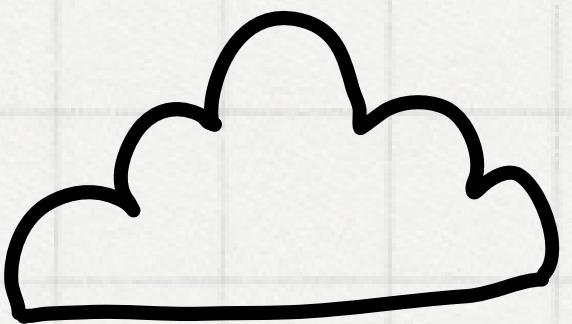
Total Revenue by Distributors



Sales by Branch Area

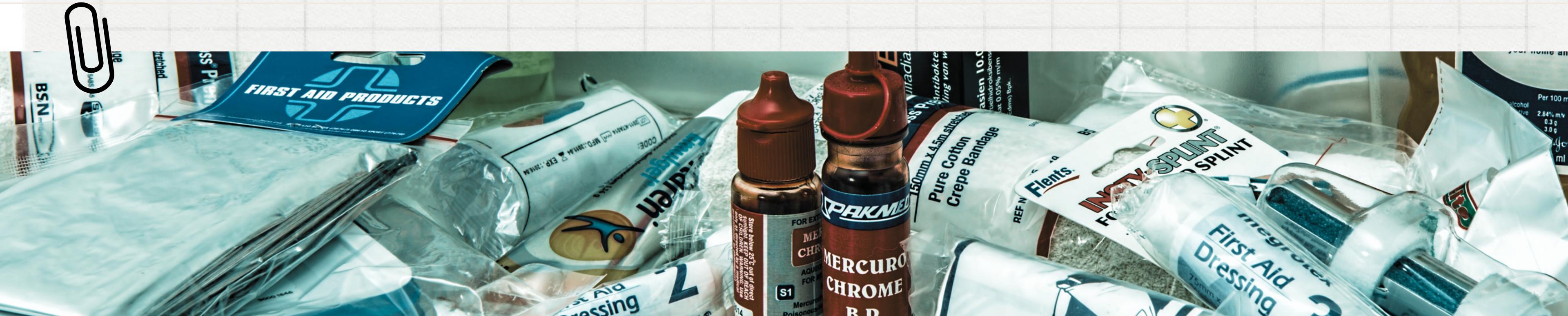


Soal 6



Dari data yang tersedia, menurut kamu untuk melengkapi analisis nya apakah diperlukan data lain juga? jika iya, sebutkan data apa yang kamu maksud dan mengapa memerlukan data tersebut

Diperlukan penggolongan kategori jenis obat, seperti obat bebas, obat bebas terbatas, obat keras, obat golongan narkotika, obat fitofarmaka, obat herbal terstandar (OHT), obat herbal (jamu). Hal ini dilakukan untuk memberikan kerangka yang terstruktur yang jelas dalam mengatur obat - obatan dan memperkuat keamanan serta meningkatkan ketepatan distribusi obat.



By Anggita Ramadhani

Thank
you!